Supplementary note to accompany

An epigenome-wide association study meta-analysis of educational attainment

1 Study overview

This meta-analysis study of educational attainment (EA) and cytosine-phosphate-guanine (CpG) methylation in human DNA was performed according to a pre-specified analysis plan. The analysis plan was publicly archived on Open Science Framework (OSF) in September 2015 (available at https://osf.io/9v3nk/), and it specified two main analyses to be conducted at the cohort level – an *epigenome-wide association study* (EWAS)[1], and an *epigenetic clock analysis*[2]. The EWAS is considered hypothesis-free as it is performed genome-wide without an expected direction of effect for individual CpG loci (often referred to as CpG probes), while the epigenetic clock analysis is hypothesis-driven since we *a priori* expect lower EA to be associated with a faster running epigenetic clock (a higher DNA methylation age). The methods of the EWAS and the epigenetic clock are further explained in **2. Epigenome-wide association study (EWAS)** and **3. Epigenetic clock analyses**.

When we designed the study we aimed to achieve a sample size of at least 10,000 individuals, which would lead to 80% power to detect an effect as small as $R^2 = 0.38\%$. This effect size is much smaller than those found in EWAS of other traits[3–5], but smaller effect sizes could possibly be expected as EA is a biologically distal environmental factor. To achieve the desired sample size we were required to pool data across cohort studies and perform a meta-analysis. Due to privacy and data sharing restrictions the EWAS and epigenetic clock analysis could only be performed locally, and subsequently meta-analysed. Cohorts could join this study by providing a signed collaboration agreement and sample descriptives during the fall of 2015. The Principal Investigator (PI) of each cohort affirmed that the results contributed to the study were based on analyses approved by the local Research Ethics Committee and/or Institutional Review Board responsible for overseeing research. All participants provided written informed consent. An overview of the 27 independent cohort studies from the 15 participating cohorts is reported in **Supplementary Table S1.1**.

In summary, the EWAS was performed with a *basic model* with covariates for age, sex and white blood-cell counts, and an *adjusted model* with additional covariates to correct for the two lifestyle factors body mass index (BMI) and smoking, which were available in all cohorts. For a full description of the control variables see **Supplementary Note 2.1.1**. Since lifestyle factors such as BMI and smoking are known to be strongly associated with both methylation and EA[6–8], we focus the presentation and follow-up analyses on the results of the adjusted model as we consider this to be more conservative.

1.1 Sample inclusion criteria

The analysis plan specified the individual inclusion criteria, limiting the analysis to individuals of European ancestry, who were at least 25 years of age at the time of assessment of EA. Individuals were to be excluded if they were cases from cohorts with a case-control study design; if they did not have successfully measured methylation or did not pass other cohort-specific standard quality control (QC) filters. Cohort-level individual and CpG probe filtering is explained in **1.3 DNA methylation measurement and cohort-level quality control (QC).** A second quality control was implemented after meta-analysis, and this procedure is described in **2.2 Description of major steps in meta-level quality-control (QC) analyses**.

The total meta-analysis sample sizes of the basic and adjusted EWAS models are respectively 10,767, and 10,317. Some individuals were excluded from the adjusted model due to missing covariates, which caused the difference in sample size.

Cohorts performed the analyses with code pre-specified in the analysis plan[9], and this code can be accessed at Open Science Framework[a]. For the EWAS the cohorts provided the following summary statistics for each probe to the meta-level analysis - CpG probe ID (markername), estimated regression coefficient (beta), standard error of the coefficient estimate, *P*-value of the coefficient estimate, and sample size. For the epigenetic clock analyses the cohorts provided the parameter estimates for the four different age acceleration measures, standard error of the parameter estimate, *P*-value of the parameter estimate, and sample size.

1.2 Phenotype definition and sample descriptive statistics

Educational attainment (EA) was harmonized across cohorts in accordance with earlier work of the Social Science Genetics Association Consortium (SSGAC)[10,11], i.e., in accordance with the ISCED 1997 classification (UNESCO)[12]. The classification consists of seven internationally comparable categories, which were subsequently translated into US years-of-schooling equivalents, which have a quantitative interpretation (**Supplementary Table S1.2**). The translated measure maintains a high level of variance in the phenotype and the cohort-specific translation is reported in **Supplementary Table S1.3,** together with the within-cohort means and standard deviations.

A summary of the 27 independent cohort studies from the 15 contributing cohorts is reported in **Supplementary Table S1.1**. The within-cohort mean age at reporting ranges ranges from 26.6 to 79.1 years, and the minimum and maximum age is respectively 18[b] and 94 years. The sample size ranges from 48 to 1,658, with an average of 399 individuals. The average EA within the cohorts ranges from 8.6 to 18.3 years of education, and the sample-size weighted average is 13.6 (SD = 3.62). Females comprise 54.1% of the meta-analysis sample.

1.3 DNA methylation measurement and cohort-level quality control (QC)

Genome-wide CpG methylation was measured in whole blood with the Illumina 450k Human Methylation chip. Since background correction and normalisation of methylation data is time-consuming and dependent on sample-specific properties, and as no method is considered overall superior, we encouraged the cohorts to perform background correction[13] and normalization[14] according to their standard QC protocols to prepare the chip data for analyses. We report the cohort-specific technical details in **Supplementary Table S1.4.**

Cohorts were recommended to implement exclusions of CpG probes and samples of individual participants using the following thresholds: probes should be excluded if the probe-detection *P*-value was greater than 0.01 in more than 5% of the individuals, and individuals should be excluded if the detection *P*-value was greater than 0.01 for more than 5% of the probes within an individual.

[a] To access the code for follow-up analyses please contact the corresponding author.

[b] The FTC and LLD cohorts were allowed to additionally include individuals slightly below 25 years of age because the FTC cohort includes on-going education in the EA measurement, and the cohorts has validated the phenotype with later cohort waves (Cohen's $\kappa = 0.82$). The LLD cohort measures EA with high enough precision to capture most of the variation even for individuals that had not yet begun higher education.

White blood cell counts, which were included as covariates to avoid confounding due to differential leukocyte cell composition across individuals, could either be measured directly or imputed based on the Houseman algorithm[15].

To summarize, we standardized the analysis protocol as much as possible, while ensuring some degree of flexibility to keep the implementation feasible for all samples.

2 Epigenome-wide association study (EWAS)

The dependent variable in the epigenome-wide association study (EWAS) was the methylation beta-value, i.e., the proportion of methylation at a CpG locus across the measured cells within an individual. The beta-value hence lies in the interval [0, 1], and has a biologically meaningful interpretation. An alternative approach is to regress the methylation as the so-called M-value, which is the log2 transformed beta-value[16]. For comparability, we used the beta-values following the methodology of the largest EWAS studies to our knowledge[3,5,6,17]. The relation between CpG methylation and the technical covariates motivates having the methylation as the dependent variable since the technical covariates reduce the error variance leading to greater statistical power to find associations with the phenotype of interest.

Each cohort study performed estimated the following regression for each CpG probe passing cohort-level quality control:

$$CpG_i = \beta_0 + \beta_1 EA_i + \boldsymbol{PC_i\gamma} + \boldsymbol{X_i\alpha} + \boldsymbol{C_i\theta} + \epsilon_i, \quad (1)$$

where CpG_i is the methylation beta-value for individual i, EA_i is the harmonized continuous measure of EA, $\boldsymbol{PC_i}$ is a vector of the first four principal components of the genetic relatedness matrix, and $\boldsymbol{X_i}$ is a vector of control variables further explained below. $\boldsymbol{C_i}$ is a vector containing study-specific controls and technical covariates (such as dummy variables for plates, hybridisation date, and batches) that were encouraged in the analysis plan and discussed below.

In accordance with the pre-specified analysis plan, probes with *P*-value less than 1×10^{-7} (the commonly-used threshold in epigenome-wide association studies[1]) were considered as epigenome-wide significant associations. We report two-tailed *P*-values throughout the paper unless otherwise specified.

2.1.1 Control variables in the adjusted model

Two epigenome-wide association analyses were performed in each cohort, a *basic model* and an *adjusted model*. The two models differ in the included covariates ($\boldsymbol{X_i}$) in accordance with the pre-specified analysis plan, where the basic model controls for age, sex, and white blood-cell counts. The adjusted model additionally includes BMI (kg/m^2), and smoking measured as a categorical variable (measured either as Ever/Never smoker or Current/Former/Never smoker, depending on data availability), together with a squared age term to account for non-linear age effects, and an interaction term between age and sex. Since lifestyle factors such as BMI and smoking are known to be strongly associated with both methylation and EA[6–8], we focus the article on the results of the adjusted model as we consider this to be more conservative.

The quadratic age term and the interaction between age and sex were added to the adjusted model, because of the known non-linear relationship between age and CpG methylation (see Bollati et al., 2009; Langevin et al., 2011; Horvath et al., 2012; Bell et al., 2013; Florath et al., 2014)[18–22], as well as the relationship between sex and CpG methylation (e.g. Boks et al., 2009; Liu et al., 2010; Zhang et al. 2011)[23–25]. Moreover, levels of educational attainment (our main explanatory variable) vary by age (birthyear) and sex (see e.g. Barro and Lee, 2001)[26]. Hence, the inclusion of these control variables is both biologically and statistically warranted.

2.1.2 Cohort-specific control variables

After discussion with the meta-analysts a few additional cohort-specific control variables were encouraged if they were deemed necessary to control for confounding relationships between EA and CpG methylation, and these cohort-specific control variables were included in $\boldsymbol{C_i}$. The EPIC and MCCS cohorts did not have genetic data available to allow inclusion of genetic PCs, so the first principal component of the methylation matrix was included instead as a stringent alternative to control for possible inflation of the test statistic due to subtle population stratification. The HBCS cohort included a dummy for childhood separation exposure, KORA F4 a dummy for World War 2, and MCSS included country of birth control variables, all of which could plausibly be correlated with both EA and methylation.

Moreover, in ALSPAC, EPIC-Breast Cancer, MCCS-Breast Cancer, and MCCS-Prostate Cancer, sex was not included because these samples consist of solely females or males. In EGCUT1 smoking was not included because all participants were non-smokers. The EPIC and MCSS samples did not control for genetic PCs because of unavailability of these variables but controlled instead for the first PC of the methylation data as a stringent alternative. The age^2 term was not included in the LBC models due to the very narrow age-range in these birth cohort samples.

2.2 Description of major steps in meta-level quality-control (QC) analyses

A stringent quality-control (QC) protocol was performed to ensure that only high-quality CpG probes were meta-analysed. All cohorts were asked to supply descriptive statistics and phenotype definitions according to the pre-specified analysis plan, and the completeness of these documents was assessed as the first step of quality control, together with examination of the uploaded EWAS summary statistics. Thereafter we applied the following probe filters.

Filters were applied to remove probes with missing *P*-value, standard error or coefficient estimate; probes not available in the probe annotation reference by Price et al. (2013)[27]; CpH probes (H = A/C/T); probes on the sex chromosomes; cross-reactive probes highlighted in a recent paper by Chen et al.[28]; and probes with a cohort-level call rate less than 95%.

Probes were annotated if they were so-called *SNP-probes*, i.e., if the probe is located on a single-nucleotide polymorphism (SNP). This was done so that any positive results could be interpreted with caution if the associated probe would be located on a SNP, as this is known to affect the methylation status of the CpG probe[27]. We however chose to keep all SNP-probes in the results, rather than removing all of them.

The output from the quality control was examined to see if any filters removed an unusual or unexpected number of probes, and two analysts independently performed and crosschecked the QC. After probe filtering, the distributions of the coefficient estimates (betas) were compared across the cohorts to identify possible outliers and birth-year effects.

2.3 Meta-analysis

Due to the differences in the mean and standard deviation of CpG methylation across cohorts we decided to perform sample-size weighted fixed-effect meta-analysis of the cohort-level EWAS summary statistics using the METAL[29] software, as fixed effect meta-analysis is robust to differences in units of measurement. Due to the variability of λ_{GC} across cohorts we applied cohort-level genomic control to deflate the association test statistic prior to meta-

analysis, equivalent to its GWAS analogue[30], to stringently control for possible population stratification that could remain even after controlling for genetic principal components in regression. In the final meta-analysis results, only probes with a meta-level sample size greater than 1,000 were considered.

2.4 Results – Epigenome-wide association study

2.4.1 Quality control results

The result of the quality-control filtering is reported in **Supplementary Table S1.5** where we also report the post-QC cohort-level genomic inflation factor, λ_{GC}, which we define here as its GWAS analogue[31], i.e., as the ratio of the median of the empirically observed distribution of the chi-square test statistic to the expected median of the null distribution.

After quality control, the remaining probes were used to produce quantile-quantile (QQ) plots for visual inspection. Most cohorts showed no sign of inflation of the association test statistics, while the cohorts LLD, MCCS lung, and RS3 displayed some inflation among probes with *P*-values above 0.05, and a few other cohorts instead displayed deflation. Inflation can be expected if a probe is associated by chance in a cohort due to the correlation between the chance finding and other probes[32], as well as if the trait is truly associated with many probes, similar to the effect on λ_{GC} caused by polygenicity in GWAS[30]. Compared to other EWAS of similar sample size[4], we also noticed some variability in the λ_{GC} between cohorts, and the average λ_{GC} of the adjusted model was 1.022 (SD = 0.175). For stringency we applied genomic control deflation of the test statistics of each cohort prior to meta-analysis[32], which is further explained in section **2.3 Meta-analysis**. After genomic control the meta-analysis λ_{GC} is 1.195 for the basic model, and 1.061 for the adjusted model.

We analysed the heterogeneity of effect-size directions across the cohort studies to detect possible outliers. We report the cohort-level direction of effects in **Supplementary Table S1.6a** and **Supplementary Table S1.7a**. To probe the validity of using OLS in the EWAS, we inspected the distribution of the beta values for the lead probes and verified that the distribution of the error term in the adjusted model of these probes was (approximately) normal.

2.4.1 EWAS results

Manhattan plots and QQ-plots of the meta-analysis of the basic model and the adjusted model are presented in **Supplementary Figure 1** and **2**. There were 37 CpG probes associated with EA at epigenome-wide significance threshold ($P < 10^{-7}$) in the basic model, the results are reported in **Supplementary Table S1.6a**, and 9 probes were found to be associated in the adjusted model (all of which were associated in the basic model). The 9 associated probes from the adjusted model are listed in **Supplementary Table S1.7a**, and we hereafter refer to these 9 as the *lead probes*.

We note that based on physical position, three lead probes are located within a 1kb window on chromosome 2, two lead probes are within 26kb on chromosome 5, and two lead probes are within 0.5kb on chromosome 7. If we prune the lead probes by keeping the strongest association within a 250kb window, then there are five “approximately independent” lead probes. Even though the independence of these associations is not certain due to their proximity, we choose to display all of the 9 associated probes as is customary in the EWAS literature.

2.4.1 Benjamini-Hochberg correction

As an alternative to the stringent epigenome-wide significance threshold ($P < 10^{-7}$), we investigated the False Discovery Rate (FDR) as is common in the EWAS literature. FDR-controlling procedures are used to control the expected proportion of rejected null hypotheses that are incorrect rejections. As our analysis plan did not specify that an FDR-correction would be applied, we present this for completeness with the EWAS literature but do not focus on the individual probes[c]. Hence, if we instead used a Benjamini-Hochberg FDR threshold of 0.05, then there would be 364 probes passing the threshold for the basic model, and 26 for the adjusted model **Supplementary Table S1.6b** and **Supplementary Table S1.7b**[33].

2.4.2 Comparison of EWAS effect-sizes with other phenotypes

In **Figure 2** and **Supplementary Table S1.8** we present the effect-size estimates, in terms of R^2, of the 50 most associated probes of our adjusted EWAS model. For comparison we contrast the results with the 50 most associated probes from recent, large-scale EWAS on alcohol consumption (n = 9,643, Liu et al., 2016)[17], BMI (n = 7,798, Mendelson et al., 2017)[3], smoking (n = 9,389, Joehanes et al., 2016)[6], and maternal smoking (n = 6,685, Joubert et al., 2016)[4], and as a comparison between EWAS and GWAS we also display the top 50 "approximately independent" SNPs from a large (n = 405,073), and recent GWAS meta-analysis of educational attainment by Okbay et al. (2016)[34].

Noticeable is the large difference between the strong effects of smoking (largest R^2 = 2.2%) and maternal smoking (largest R^2 = 11.6%), in comparison with our EWAS results (largest R^2 = 0.69%). Also the effects of BMI on methylation are consistently stronger than those of EA, while the effects of alcohol consumption are more similar, but still consistently larger than those of EA. There are also large differences between EWAS and GWAS of EA, where the SNP effects are an order of magnitude smaller than the CpG probe effect sizes.

[c] Many of the follow-up analyses include probes from the adjusted model with P-value less than 1×10^{-4} or 1×10^{-5}, and these analyses therefore include more information than just the associations of the 9 lead probes.

3 Epigenetic clock analyses

The association between EA and biological ageing was examined with an epigenetic clock analysis[2]. *A priori*, we hypothesised that lower EA would be associated with a faster rate of ageing, resulting in a faster running epigenetic clock. The epigenetic age was assessed according to the specified analysis protocol by entering raw CpG beta value data into the online Horvath calculator[d]. The "normalize data" and "advanced analysis for Blood Data" options were selected. The following biological clock variables were selected from the calculator's output for downstream analysis:

1) Horvath age acceleration residuals (Clock 1), which are the residuals from the regression of chronological age on Horvath age.

2) White blood cell count adjusted Horvath age acceleration (Clock 2), which is the residual from (1) after additional covariate adjustment for imputed white blood cell counts.

3) White blood cell count adjusted Hannum age acceleration (Clock 3), which is the same as (2) but with the Hannum age prediction in place of the Horvath prediction.

4). Cell-count enriched Hannum age acceleration (Clock 4), which is the basic Hannum predictor plus a weighted average of ageing associated cell counts. This index has been found to give the strongest association between epigenetic age acceleration and mortality[35].

Linear regression was used to test the associations between education and the four selected measures of DNA methylation age acceleration. Two models were considered for each age acceleration measure: adjustment for age and sex (*basic age acceleration model*), and additional control variables for body mass index and smoking (*adjusted age acceleration model*). Sample-size weighted meta-analyses of the results were performed across cohorts.

3.1 Results – Epigenetic clock analyses

The results of the epigenetic clock analyses are presented in **Supplementary Table S1.9** and **Figure 4.** For the epigenetic clock analyses, the cohorts BSGS (only included in the basic model because of absent covariates for the adjusted model), FINRISK, and MCCS could not contribute. Therefore, the sample size for the basic age acceleration model was 8,173, and was 7,691 for the adjusted age acceleration model. We did not observe evidence of an association between EA and the epigenetic clock for Clock 1 (Horvath age acceleration residuals), Clock 2 (White blood cell count adjusted Horvath age acceleration), and Clock 3 (White blood cell count adjusted Hannum age acceleration). We found evidence for an association between EA and Clock 4 (Cell-count enriched Hannum age acceleration, $P = 3.51\times10^{-6}$ in the basic age acceleration model and $P = 4.51\times10^{-4}$ in the adjusted age acceleration model). We note that this latter measure (Hannum age acceleration with dynamic weights) is also the most predictive epigenetic clock measure of mortality[35]. Due to the inconsistent results across the different clock variables and the small estimated effect sizes, we refrain from drawing strong conclusion based on the epigenetic clock results.

[d] The Horvath calculator can be accessed at https://labs.genetics.ucla.edu/horvath/dnamage/.

4 Investigation of possible confounding with tobacco smoking

The adjusted EWAS model included a categorical control variable for tobacco smoking (categorised as current, former or never smoker). Due to the discrete nature of the variable and the measurement error that follows by not measuring tobacco smoking as a continuous variable, we believe that the control variable might not have completely controlled for the exact smoking exposure. Unfortunately, controlling for smoking with a categorical variable is the common approach due to the lack of a more precise smoking measure in most cohort studies. As a result, smoking could potentially bias the regression coefficient of EA because of the known negative correlation between smoking and EA[36], and the strong association between smoking and methylation found in a large number of studies[4,6]. Therefore, we performed a literature review using PubMed to see if any of the lead probes from the adjusted model have been associated with smoking in previous studies.

The literature review was performed February 24, 2016 by searching for the term "smoking" together with each of the adjusted model's 9 lead probes separately. The search resulted in 30 eligible studies or smaller meta-analyses (n < 1,800) on CpG methylation and different forms of smoking exposure, one systematic review by Gao et al. (2015)[37], and two larger meta-analysis by Joehanes et al. (2016, n = 9,389)[6] and Joubert et al. (2016, n = 6,685)[4] that include many of the individual studies. The EWAS meta-analysis by Joehanes et al., 2016[6], comprising of several of the 30 individual studies, was published after the literature review had been performed and we added this study to the literature review post-hoc due to its relevance and large sample size. The result of the literature review is presented in **Supplementary Table S.1.10**.

The studies could be categorised as investigating either the relationship between a person's own smoking and methylation, or the association between newborn's methylation and maternal smoking, both at birth and at later stages in the life of the offspring. Most of the individual studies have relatively small sample sizes (smallest sample size consists of only 21 monozygotic twin pairs, and the average sample size is 471), and we note that many of the individual studies on personal smoking are meta-analysed in Joehanes et al. (2016), and most of the individual studies on maternal smoking are meta-analysed in Joubert et al. (2016).

We found that all of the adjusted model's 9 lead probes have been associated with smoking in at least one previous study, and in **Supplementary Note 4.3** we contrast the EWAS effect sizes estimates for EA with the effect size estimates from the two meta-analyses studies. However, not all of these previous studies controlled for EA, and the associations between smoking and methylation could therefore be biased due to the non-inclusion of EA as a control variable, and this makes it complicated to draw definite conclusions of the magnitude of confounding. Also, the sample sizes of many of the individual studies are small and likely to contain a higher number of false positives than the meta-analyses.

In light of this finding we interpret our EWAS findings carefully, and the literature review motivated the *post-hoc* sensitivity analyses in the subsample of people that never smoked presented in the next section.

4.1 Robustness of EWAS results in the never-smoker subsample

To further investigate the possible confounding effect of smoking on the association between EA and CpG methylation, all cohorts reran the EWAS in the subsample of individuals that reported as 'never smoker'. We refer to this subsample as the *never smokers*. The full EWAS sample consists of the *ever smokers* (i.e., the individuals who answered that they were ever,

current, or former smokers), and the never smokers. This approach would give us confidence in the associations of the adjusted model if the lead probes were found to be associated with EA in the never-smoker subsample. We note that this sample could be biased in other ways when selected on this specific variable, and the statistical power will be lower due to the decrease in sample size, and we therefore do not perform any analyses of any other probes than the 9 lead probes in this subset.

The meta-analysis sample size of the never-smoker subsample was 5,175, and with the effect size estimates from the adjusted model we calculated the power to find the effects in the never-smoker subsample. The expected power ranged from 6.4% to 74% (**Supplementary Table S1.7a**), at *P*-value < 10^{-7}, and the expected number of replications was 1.93 given the expected power. If a more liberal Bonferroni-corrected threshold would be used, such as 0.05 divided by 9, then the power in the never smokers ranged from 81% to 100%. Hence, we expected considerable statistical power to replicate the lead probes of the adjusted model in the never-smoker subsample assuming the full sample effect size estimates.

After quality control, we performed a meta-analysis of the re-estimation in the never-smoker subsample, and the results are displayed in **Supplementary Table S1.7a**. None of the 9 lead probes were significant at the stringent epigenome-wide threshold of $P < 10^{-7}$, while 2 probes, cg12803068 and cg22132788, were estimated with strong associations even though the sample size was halved (with respective *P*-values 1.48×10^{-4} and 4.35×10^{-4}).

4.2 Joint test of no association in the never-smoker subsample

We thereafter performed a joint test of the null hypothesis that the regression coefficients of the lead probes were all equal to zero, with the sample-size weighted *Z*-statistic estimated in the never smokers. Since we have no a priori hypothesis of a direction of effect for different CpG probes, and to avoid that positive and negative *Z*-statistics cancel out, we performed this test on the absolute value of the *Z*-statistic with the following weighting

$$Z_{Combined} = \frac{\sum_{i=1}^{k} w_i Z_i}{\sqrt{\sum_{i=1}^{k} w_i^2}} \quad (2)$$

where

$$w_i = \sqrt{N_i}$$

The right-tailed *P*-value of the joint test of no association was 2.18×10^{-11}. As a robustness check we also performed this test while pruning the lead probes so that only the strongest association within 250kb was kept, and a right-tailed *P*-value of 4.91×10^{-6} based on 5 probes as the result. We interpret this as evidence that at least one probe has an estimated effect different from zero in the meta-analysis of the never smokers (however see **Supplementary Note 4.4** for more discussion of maternal smoking as a confounding factor).

4.3 Effect size comparison ever- versus never smokers

We thereafter investigated the change in effect size estimates between the ever smokers and the never smokers. To be able to compare the differences in effect sizes and the precision of the estimates (SE), we used inverse-variance weighted meta-analysis to retain meta-analytical betas and standard errors for the 9 lead probes. We thereafter derived the effect sizes and standard errors of the ever smokers by assuming that the effect size estimates in the full

sample is a weighted average of the effect size estimates in the ever smokers and never smokers, as the two groups are mutually exclusive and the union between the two is the full EWAS sample.

The result of the effect size comparison is presented in **Figure 3**. We aligned the effect sizes to the first quadrant by taking their absolute values so that all effect sizes were positive for the comparison. If the effect sizes would be similar across the subsamples we would expect them to align along the 45-degree line, and if confounding would be driving the effect size estimates then we would expect a cluster above the 45-degree line as the effect size estimates in the ever smokers would be greater than those in the never smokers.

By visual inspection we observed a cluster of probes with larger effect size estimates in the ever smokers than in the never smokers (i.e., those above the 45 degree line). We also observed a second cluster of effect size estimates with similar estimates in both subsamples (i.e., those along the 45 degree line). The 2 lead probes that could potentially be associated in the never smokers i.e., cg12803068 and cg22132788, have slightly larger effect size estimates in the ever smokers than in the never smokers, however these larger estimates lie within the confidence intervals of the estimates in the ever smokers. The other 7 probes have effect size that are more than 50% smaller in the never smokers than in the ever smokers, and all of these effect size estimates are outside the confidence intervals of the estimates among ever smokers.

4.4 Lookup of probes in EWAS meta-analysis on smoking and maternal smoking during pregnancy

In **Figure 3** and **Supplementary Table S1.11** we present the effect-size estimates, in terms of R^2, of the 44 probes of the adjusted model with P-value less than 1×10^{-5} (including the 9 lead probes), ordered descendingly on R^2, and in contrast we display these probes' effect sizes when re-estimated in the never-smoker subsample, as well as the effect size of these probes as reported in the recent EWAS meta-analyses on smoking by Joehanes et al. (2016)[6], and maternal smoking by Joubert et al. (2016)[4]. We report the effect sizes for smoking and maternal smoking if a probe was significantly associated in these studies at FDR < 0.05, as those results were publicly available.

Noticeable is that the effect sizes of smoking and maternal smoking are many times larger for most probes compared to the effect of EA. Secondly, the probes cg12803068 and cg22132788 that were estimated with low P-values in the never smokers, and the closest gene; *MYO1G*, has been estimated as persistently affected by maternal smoking in new-born's as well as later in the life of the offspring[4,38], and the effect of maternal smoking on these two probes is extreme; the R^2s are respectively 8.34% and and 6.33%. We have no possibility to control for maternal smoking and we can therefore not distinguish between the hypothesis that these probes have some true association with EA and the hypothesis that their apparent association with EA is entirely driven by more maternal smoking during pregnancy among lower-EA individuals. The extreme effect size of smoking and maternal smoking substantially weakens the possibility that these probes would be directly affected by EA, rather than indirectly by the confounding. We also cannot rule out that the probes' association with EA is driven by second-hand smoke exposure, which could also be correlated with EA, and is not controlled for in regression. The lookup strongly suggests that smoking and maternal smoking remain potential confounding factors for the probe associations with EA, even in the subsample of individuals who are self-reported never-smokers.

5 Enrichment analyses

5.1 Enrichment analysis of GWAS of smoking and EA

We performed an enrichment analysis to investigate if the genetic loci, in proximity with the probes found to be associated in the adjusted EWAS model (at $P < 1\times10^{-4}$), also contain SNPs enriched for EA and smoking. For EA we used the GWAS summary-statistics from a large (n = 405,073), recent meta-analysis by Okbay et al. (2016)[34]. For smoking we used a meta-analysis combining GWAS results from the UK Biobank together with the publically available summary statistics from the Tobacco, Alcohol, and Genetics Consortium (TAG, total n = 186,102)[39], and this meta-analysis is described in detail in Karlsson Linnér et al. (Manuscript in preparation)[40].

The enrichment analysis was performed with the 179 EWAS probes from the adjusted model with *P*-value less than 1×10^{-4}. We pruned the probes so that the probe with the lowest *P*-value in a locus was selected as the lead probe, and all probes within 250kb from the lead probe was clumped with the strongest association. This resulted in 141 "approximately independent" lead probes (k). For each of the phenotypes, we extracted the closest SNP available in the GWAS summary statistics based on the physical position of the k lead probes. Using the reference panel 1000 Genomes phase 3 (October 2014 haplotype release version 5), we extracted the SNPs in strong LD (r^2 >= 0.8) with the k SNPs, and then we averaged the absolute value of their *Z*-statistics of the association with the phenotype, as well as the minor allele frequency (MAF) of the SNPs in strong LD. This leads to a distribution of k average *Z*-statistics and average MAF.

From the summary statistics we extracted a set of 1,000 random SNPs for each of the k SNPs, matched on the average MAF (+/- 1pp), and for the matched SNPs we also averaged the absolute value of the *Z*-statistics across the SNPs in strong LD, just as for the 141 "first-stage" SNPs. This leads to a distribution of $k\times1{,}000$ average absolute *Z*-statistics. We ordered the *Z*-statistics and performed a test of joint enrichment with the non-parametric Mann-Whitney test of the null hypothesis that the association test statistic of the k SNPs are drawn from the same distribution as those of the 141,000 MAF-matched SNPs.

5.1.1 Results – Enrichment analysis of GWAS of smoking and EA

For EA, the average absolute value of the *Z*-statistics of the k SNPs is 1.199, the median 0.956, and the average of the two-tailed *P*-values is 0.377. For the matched SNPs the average absolute value of the *Z*-statistics is 1.101, and the median is 0.889, and the average of the two-tailed *P*-values is 0.406. The result of the Mann-Whitney test does not reject the null hypothesis of no enrichment for EA (two-tailed *P*-value = 0.206).

For smoking the average absolute value of the *Z*-statistics of the k SNPs is 0.953, the median 0.846, and the average of the two-tailed *P*-values is 0.437. For the matched SNPs the average absolute value of the *Z*-statistics is 0.921, and the median is 0.760, and the average of the two-tailed *P*-values is 0.453. The result of the Mann-Whitney test does not reject the null hypothesis of no enrichment for smoking (two-tailed *P*-value = 0.504).

5.1.2 Enrichment of individual SNPs

We continued by examining if any of the average absolute *Z*-statistics of the 141 "approximately independent" lead probes (k) were associated with the phenotypes. No loci were found to be associated with smoking (smallest two-tailed *P*-value = 0.001), while one

SNP, rs9956387, is associated with EA, two-tailed *P*-value of the average *Z*-statistic = 4.39×10^{-8}. The CpG probe related to the SNP, cg17939805, has *P*-value = 7.02×10^{-5} in the adjusted EWAS model results, and the distance between the probe and the SNP is 607 base pairs.

5.2 Enrichment within methylation density regions (HIL)

We clustered the meta-analysis results into the four methylation density categories ("HIL") defined in Price et al., (2013)[27], using the annotation file from that study. Each probe could be categorized as located in a high-density CpG island (HC), an intermediate-density CpG island (IC), a region of intermediate-density CpG island bordering HCs (ICshore), or a non-island (LC). We investigated the enrichment for association of probes in these four categories. The results can be found in **Supplementary Table S1.12.**

The number of probes with $P < 10^{-7}$ is more or less proportional to the total number of probes in every category. The mean absolute *Z*-statistic is similar across categories, suggesting no difference in enrichment across categories. The enrichment statistic is defined as:

$$Z = \frac{\frac{1}{S}\sum_{s=1}^{S} z_s^2 - 1}{\sqrt{s/S}} \quad (3)$$

where *s* indexes the number of probes, and z_s^2 is the squared *z*-statistic from the regression of probe *s* on educational attainment (see also section 11 of the Supporting Information in Rietveld et al., 2014)[41]. This squared *z*-statistic captures the strength of the association between educational attainment and probe *s*, while ignoring the sign of the association. Under the null hypothesis, each $z_s \sim N(0,1)$, and thus $z_s^2 \sim \chi^2(1)$, which has mean 1 and variance 2. Therefore, under the null:

$$E(Z) = 0, var(Z) = \frac{\left(1/S\right)^2 S var(Z_s^2)}{2/S} = 1 \quad (4)$$

We calculate a *P*-value for the test of whether the realized value of the test statistic, $Z = z$, differs from zero using the inverse cumulative distribution function of the standard normal distribution. The results indicate that there is enrichment for association in all four HIL categories.

6 Prediction of EA with polygenic methylation score

To test the predictive accuracy of our EWAS findings we constructed polygenic methylation scores (PGMSs) to perform out-of-sample polygenic prediction of EA. The prediction was performed in three independent cohorts studies, the Lothian Birth Cohort 1936 (LBC1936, n = 917), Rotterdam Study BIOS (RS-BIOS, n = 671), and Rotterdam Study 3 (RS3, n = 729). For each of the prediction cohorts we created a new EWAS meta-analysis withholding the respective prediction cohort study to avoid overfitting[42]. The effect sizes (in terms of Z-statistics) from the respective holdout meta-analysis were used as weights when constructing the PGMS. The Z-statistics were used instead of the EWAS coefficients because CpG methylation is the dependent variable in the EWAS regression.

For each prediction sample we created four PGMS using two P-value thresholds $P < 10^{-5}$ and $P < 10^{-7}$, and the estimates from the basic and adjusted models. The PGMS was defined as the weighted sum of the Q CpG probes:

$$\widehat{M}_i = \sum_{j=1}^{q} \hat{Z}_j \, \mathrm{B}_{ij} \,, \tag{5}$$

where $\widehat{M}_i$ denotes the methylation score of individual i, $\hat{Z}_j$ is the estimated Z-statistic for CpG probe j, and B_{ij} is the methylation beta-value for individual i at CpG probe j.

A recent study showed that the phenotypic prediction can be improved if the PGMS is combined with a single-nucleotide polymorphism polygenic score (SNP PGS)[43]. We therefore included a SNP PGS constructed with the effect size estimates from a recent large-scale GWAS on EA (n = 405,073)[11], and the prediction cohorts were excluded from the GWAS meta-analysis to avoid overfitting. The SNP PGS was defined as the weighted sum of the T directly genotyped SNPs:

$$\hat{S}_i = \sum_{j=1}^{t} \hat{\beta}_j \, g_{ij} \,, \tag{6}$$

where $\hat{S}_i$ denotes the polygenic score of individual i, $\hat{\beta}_j$ is the estimated additive effect size of the effect-coded allele at SNP j, and g_{ij} is the genotype of individual i at SNP j (coded as having 0, 1 or 2 instances of the effect-coded allele)[44].

For each prediction cohort study we performed an OLS regression with EA as the dependent variable, and age and sex as control variables. The predictive power of the PGMS was evaluated as the incremental R^2 of adding the PGMS to the regression model, i.e., the difference in R^2 between the regression model with only the SNP PGS and the covariates, and the regression model with the PGMS together with the SNP PGS and the covariates. The incremental R^2 of the interaction term (between the PGMS and SNP PGS) was evaluated as the difference in R^2 between the regression model with the PGMS and the SNP PGS as additive main effects, together with the covariates, and the same regression model with the interaction term added.

To estimate 95% confidence intervals for the incremental R^2 we performed bootstraps with 1,000 samples with the percentile interval method, and this was done with the 'boot' package in R[9,45,46]. Finally, the incremental R^2 was meta-analysed across the three prediction cohorts (LBC, RS-BIOS, and RS3) as the sample-size weighted incremental R^2.

6.1.1 ALSPAC prediction of educational achievement

A further prediction analysis, using the same PGMS prediction method as described in **Supplementary Note 6**, was carried out in the ARIES substudy (n = 678) of the ALSPAC cohort[47,38]. Cord-blood methylation was assessed in newborn children and EA was measured using four sets of standardised Key Stage[48] school grades from primary level through to high school (educational achievement was evaluated at ages 7-16 years), and the Key Stage educational achievement test scores are part of the state education system in the UK[48]. The cord blood signatures were collected prior to any educational exposure and may help identify if methylation differences might be a cause or consequence of EA. For each of the Key Stages 1 to 4, an average score across all school subjects was derived for each child. The predictive accuracy of the PGMS, built using the association Z-statistic from a meta-analysis excluding the ALSPAC cohort, was assessed as the incremental adjusted R^2 of adding the score to the OLS regression model, with sex and age at assessment as control variables. Secondly, we tested if these associations were robust to the inclusion of maternal smoking as an additional control variable. Finally, we performed the prediction in a restricted sample using data from children of non-smokers only.

6.2 Results – Prediction of EA with polygenic methylation score

The result of the out-of-sample prediction with the polygenic methylation score (PGMS) is presented in **Figure 5** and **Supplementary Table S1.13a**. The total meta-analysis sample size of the prediction across the three holdout-cohorts LBC1936, RS-BIOS, and RS3 was 2,317. Based on the basic model parameter estimates and P-value threshold 1×10^{-5}, the weighted average incremental R^2 was 1.73%, and 1.80% with P-value threshold 1×10^{-7}.

The prediction results based on the adjusted model parameter estimates were consistent with those of the basic model. The weighted average incremental R^2 was 2.00% at $P < 1\times10^{-5}$, and 1.43% at $P < 1\times10^{-7}$. Notably, the predictive power at $P < 1\times10^{-7}$ was achieved with a PGMS of only 8 probes which could explain why the PGMS at $P < 1\times10^{-5}$ had greater explanatory power, even though this approach introduces more noise due to the inclusion probes that are less strongly associated with EA.

6.2.1 Results – Interaction effect between PGMS and SNP PGS

When the interaction between the PGMS and the SNP PGS was added to the model, the R^2 additionally increased in the range from 0.11% to 0.30% depending on the P-value threshold and EWAS weights. The interaction was generally significant at $P < 0.05$, except for the basic model at $P < 1\times10^{-5}$. This suggests that there is relatively weak evidence for an interaction between the PGMS and the SNP PGS in relation to EA.

6.2.2 Results – Prediction in the never-smoker subsample

Given our concerns that smoking is a confounding factor driving the associations between methylation and EA, we also considered the inclusion of an additional PGMS in the prediction. The additional PGMS was calculated with weights (Z-statistics) from an EWAS on smoking by Zeilinger et al. (2013)[49], using the 187 probes that were identified at epigenome-wide significance ($P < 1\times10^{-7}$) and then successfully replicated. The smoking PGMS was created in three cohorts (RS3, RS-BIOS, LBC1936), and we calculated Pearson correlations between the PGMS of smoking and the PGMSs for EA constructed from the lead probes (i.e., those associated with EA at significance threshold $P < 1\times10^{-7}$) in either our basic

or adjusted model. All correlations were negative across all the EA PGMS, no matter if the scores were created with weights from the basic or adjusted model. That is, for the EA PGMS from our basic model, we find a correlation with the smoking PGMS of -0.96 in RS3, -0.94 in RS-BIOS, and -0.93 in LBC1936. For the EA PGMS from our adjusted model, the correlations are -0.90, -0.89, and -0.91, respectively. The correlations were in the expected direction, where increased EA is linked to less smoking. This strongly suggests that smoking status confounded the EWAS associations with EA. Because of strong multicollinearity we could not include the smoking PGMS to the prediction models.

We also performed the prediction in the LBC1936, RS-BIOS, and RS3 cohorts in the subsample of never smokers with the four PGMSs constructed with the parameter estimates of the full EWAS sample. This subset had a total sample size of 876 individuals across the cohorts. The results are presented in **Supplementary Table S1.13b**.

When the PGMS was constructed with probes estimated with $P < 1\times10^{-5}$, then the methylation score predicted 0.64% of EA with the basic model estimates ($P = 0.018$), and 0.93% when constructed with the adjusted model effect sizes ($P = 0.005$). No PGMS was significant when constructed at $P < 1\times10^{-7}$. The marginal significance could be explained by the reduced power due to the smaller sample (876 never smokers compared to 2,317 individuals in the full sample). No interaction effect between the PGMS and SNP PGS was found in the never-smoker subsample, as expected due to the reduced power.

6.2.3 Description of the ALSPAC cohort

ALSPAC[50] recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992. 14,541 is the initial number of pregnancies for which the mother enrolled in the ALSPAC study and had either returned at least one questionnaire or attended a "Children in Focus" clinic by 19/07/99. Of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above.

The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 18 is 706 (452 and 254 recruited during Phases II and III respectively), resulting in an additional 713 children being enrolled.

The total sample size for analyses using any data collected after the age of seven is therefore 15,247 pregnancies, resulting in 15,458 fetuses. Of this total sample of 15,458 fetuses, 14,775 were live births and 14,701 were alive at 1 year of age.

A 10% sample of the ALSPAC cohort, known as the Children in Focus (CiF) group, attended clinics at the University of Bristol at various time intervals between 4 to 61 months of age. The CiF group were chosen at random from the last 6 months of ALSPAC births (1,432 families attended at least one clinic). Excluded were those mothers who had moved out of the area or were lost to follow-up, and those partaking in another study of infant development in Avon.

6.2.1 Results – ALSPAC prediction

The only methylation score that reached nominal significance in the prediction of educational achievement test scores (Key Stage 1–4), was the prediction of Key Stage 4 with a PGMS constructed from probes with $P < 1\times10^{-7}$. The score explains 0.62% of the variance in school performance at Key Stage 4 ($P = 0.024$), and the age at assessment at this Key Stage is 15-16 years. We present these results in **Supplementary Table 1.13c**. The predictive power attenuates completely upon adjustment for maternal smoking status, as well as in the sub-group analysis of mothers who reported to have never smoked (incremental adjusted R^2 = 0.14%, $P = 0.17$). We draw two conclusions from these results from the child sample. First, they reinforce the concern that maternal smoking is a major confounding factor for all our probe associations with EA. Second, they suggest that any true methylation-EA associations are unlikely to be driven by a causal effect of methylation on EA.

7 Correlation of EWAS associations with tissue-specific methylation

The NIH Roadmap Epigenomics Consortium recently made an impressive analysis and categorization of a multitude of different epigenetic marks, across 111 different tissues[51]. The data is publicly archived, and we harnessed their tissue-specific methylation data to answer the question of whether our EWAS associations are correlated with any tissue-specific DNA methylation. We hypothesize that EWAS associations for a given phenotype would be more likely to be located at loci that are differentially methylated in tissues relevant for the phenotype or endophenotypes. E.g., if our EWAS associations would be correlated with the tissue-specific methylation of brain tissues, that would increase the credibility of the associations, and improve the biological interpretation.

Genome-wide methylation data was available for three kinds of CpG methylation measurements: whole-genome bisulfite sequencing (WGBS), reduced representation bisulfite sequencing (RRBS), and mCRF (a method combining sequencing data from the MeDIP-seq and MRE-seq methods). Methylation was measured as the beta value, ranging from 0 to 1, and it was available for 37 tissues measured with WGBS, 49 tissues with RRBS, and for 16 tissues with mCRF. Only eight tissues were available for two or more methylation measurements.

We pruned the probes of the adjusted EWAS model with *P*-value less than 1×10^{-4} using a window of 250kb, as described in **5.1 Enrichment analysis of GWAS of smoking and EA**, which resulted in 141 "approximately independent" probes. Based on the physical location of the 141 pruned probes we extracted the beta-value for WGBS, RRBS, and mCRF. The beta-values were first converted to M-values[e], and thereafter we calculated standardised tissue-specific deviations ($Z_{CpG,t}$) from the cross-tissue average methylation ($\overline{Mval}_{CpG}$) for each of the 141 CpG loci as:

(7)
$$Z_{CpG,t} = \frac{Mval_{CpG,t} - \overline{Mval}_{CpG}}{s.d(Mval_{CpG})}$$

where $Z_{CpG,t}$ is the tissue-specific deviation for tissue *t,* at locus *CpG.* This procedure was performed within each methylation measurement. Hence, in the case that a tissue had no missing CpG loci, there were 141 tissue-specific deviations; each corresponding to one of the loci identified in the association results of the adjusted EWAS model. For RRBS, there were many missing values across all 49 tissues, and the maximum number of loci available was 38 out of 141. We therefore excluded the RRBS measurement from further analysis, while both WGBS and mCRF had close to 141 overlapping loci for all tissues (**Supplementary Table 1.14**).

For WGBS and mCRF we calculated correlations with the tissue-specific *Z*-statistics and the EWAS association *Z*-statistics of the pruned probes (from the adjusted EWAS model). Bonferroni correction for the number of tissues within each methylation measurement was performed.

[e] As the consortium methylation data was only available with two decimals precision, we imputed 0's and 1's with 0.005 and 0.995 to avoid infinite M-values.

7.1 Results – Correlation of associations with tissue-specific methylation

Genome-wide, the M-values were all highly positively correlated across the tissues, both for WGBS (*r* ranged between [0.728-0.967]) and mCRF (*r* ranged between [0.758-0.925]). The correlations of the standardised deviations were lower (ranging between [–0.253-0.498] for WGBS, and [–0.175-0.264] for mCRF). Hence, the level of methylation at each locus was very similar across the tissues, while the tissue-specific deviations (that can be both positive and negative) relative to the tissue average were less correlated, and some correlations were also negative for some tissue pairs, indicating an inverse relationship relative to the average cross-tissue methylation for these tissue pairs.

The results of the correlations between the tissue-specific methylation and the *Z*-statistics of the pruned probes are reported in **Supplementary Table 1.14** and **Figure 6**, and for some tissues there were significant correlations between the tissue-specific methylation and the EWAS associations.

For WGBS, the correlations were in the range [–0.393, 0.509], and three correlations remained significant after Bonferroni-correction for the number of tissues (i.e., $P < 0.05 / 37$). The positive Bonferroni-significant correlations were *primary haematopoietic stem cells G-CSF-mobilized female* ($r = 0.509$; Roadmap ID E050, category: haematopoietic stem cells), and *thymus* ($r = 0.381$; Roadmap ID E112, category: Thymus). The negative Bonferroni-significant correlation was *IMR90 fetal lung fibroblast* ($r = –0.393$; Roadmap ID E017, category: IMR90). It should be noted that the *primary haematopoietic stem cells G-CSF-mobilized* tissue was not available in a male sample.

There were six WGBS correlations that reached marginal significance ($P < 0.05$): *right ventricle* ($r = 0.19$; Roadmap ID E105, category: Heart), *ES-UCSF4 cells* ($r = 0.185$; Roadmap ID E024, category: Embryonic stem cell), and the negative correlations were *oesophagus* ($r = –0.234$; Roadmap ID E079, category: Digestive), *HUES64 derived CD56+ mesoderm* ($r = –0.195$; Roadmap ID E013, category: Embryonic stem cell derivatives), *small intestine* ($r = –0.194$; Roadmap ID E109, category: Digestive), and *HUES64 derived CD184+ endoderm* ($r = –0.194$; Roadmap ID E011, category: Embryonic stem cell derivatives).

The correlations with the mCRF methylation measurements were in the range [–0.234, 0.162], and there were no Bonferroni-significant correlations (i.e., $P < 0.05 / 16$). There was one negative correlation with $P < 0.05$: the tissue *Breast vHMEC mammary epithelial* ($r = –0.234$; Roadmap ID E028, category: Epithelial).

There are two tissues categorized as brain tissues in the WGBS methylation data, *brain germinal matrix* ($r = 0.044$, Roadmap ID E070, *P*-value = 0.607) and *brain hippocampus middle* ($r = 0.115$, Roadmap ID E071, *P*-value = 0.175). The absolute correlation of brain hippocampus middle is just above the average of the absolute correlations ($\overline{|r|}=0.105$) across the 37 WGBS tissues, and we therefore do not interpret this as evidence in favor of differential methylation in brain tissues (i.e., a plausible endophenotype of EA) at the CpG loci found in the EWAS, compared to the other tissues.

The mCRF data contains two brain tissues, *fetal male brain* ($r = 0.162$, Roadmap ID E081, *P*-value = 0.057), and *fetal female brain* ($r = 0.037$, Roadmap ID E082, *P*-value = 0.66), and the fetal male brain correlation is the second largest absolute correlation found for the mCRF measurement, with a *P*-value close to marginal significance. However, the correlation is not as strong as the other correlations found with WGBS, and it is inconsistent with the

correlation in fetal female brain. We therefore did not interpret this in favor of our EWAS findings being correlated to differential methylation in brain.

The strongest correlations were found for haematopoietic stem cells ($r = 0.509$), which is the direct source of all other blood cells, and for fetal lung fibroblasts ($r = –0.393$). In general, the significant correlations are hard to interpret as they are spread across multiple cell types and tissues. We interpret the lack of correlation with tissues plausibly related to EA (such as brain tissues) as supporting the hypothesis that the EWAS results are driven by confounding factors rather than by a true association with EA.

As a robustness check we performed the above analysis while pruning the probes with *P*-value less than 1×10^{-2} instead of 1×10^{-4}, and the correlations were stable when compared across the *P*-value thresholds. However, when the analysis was performed with all ~450,000 probes, then the correlations attenuated to ~0 for all tissues. We think that including more probes increases the power until too many non-associated probes are included in the correlation, and the non-associated probes attenuate the correlations towards zero past some optimal *P*-value threshold with optimal signal-to-noise ratio.

8 Expression analysis

Using the GTEx expression data[52] and the webtool 'Functional mapping and annotation of genetic associations' (FUMA)[53], we clustered the expression of the 29 genes physically closest to the CpG probes of the adjusted model (with $P < 1\times10^{-5}$). Normalization of Reads Per Kilo base per Million (RPKM) was performed as zero-mean normalization with a $\log_2$ transformation of the RPKM, with pseudocount 1, and it was performed per gene. Average normalized expression values were computed per 53 specific tissue types. Genes and tissues were clustered by hierarchical clustering using pairwise Euclidian distance.

The result of the expression analysis is displayed in **Supplementary Figure 4**. By visual examination we observed one noticeable cluster of 10 genes[f] where each gene individually is more highly expressed in the brain tissues compared to all other tissues. This set of genes could possibly be related to EA due to the cognitive nature of the phenotype, however we interpret this as merely suggestive evidence, and the gene sets (containing 10 and 29 genes) are too small for a well-powered test for differential brain tissue expression. A second cluster of 19 genes[g] was observed that are more highly expressed in a wide variety of tissues including spleen, lung, thyroid, among many others, than compared to the brain tissues. Tentatively, one could expect this pattern if the genes are involved in the toxicological adverse effects of smoking in different tissues. Overall, the expression analysis did not support the hypothesis that our EWAS findings are immediately related to brain function or cognition.

[f] The genes that form a cluster with higher expression in brain tissues, compared to all other tissues, are *KCNH3, CLSTN1, YWHAQ, LINC00277, C10orf107, RNF220, MKRN3, GAL3ST3, PFZD7,* and *FLRT1.*

[g] The genes that form a cluster with higher expression in a wide variety of tissues, compared to the brain tissues, are *DAPL1, WT1, MYO15B, MYO1G, STK40, ASCL2, AHRR, DOT1L, SSR3, POLK, SP3, ARID1B, ZNF267, ZNF808, C1D, EXOC2, F2RL3, CDCP1,* and *IER3.*

9 methQTL analysis

It has been shown that genetic variants also explain variation in CpG levels[1] in addition to environmental influences on DNA methylation. EWAS probes under genetic influence may help us understand the direction of association between CpGs and outcomes, but it can also be a confounding factor. SNPs affecting the level of methylation, usually referred to as methylation quantitative trait loci (methQTL)[1], sometimes have effect sizes that can be found in samples of less than a thousand individuals[54,55], which is much less than what is necessary for most GWAS of e.g. behavioral phenotypes[34]. Therefore, as is customary in the EWAS literature, we performed a GWAS for each of the 9 lead probes to investigate if any SNPs were associated with the level of methylation. The genome-wide association analysis was performed in the LBC1936 (n = 918) and RS3 (n = 731) cohorts that estimated the following GWAS regression equation for each autosomal bi-allelic SNP and lead probe:

$$Y_i = \beta_0 + \beta_1 SNP_i + \boldsymbol{PC_i\gamma} + \boldsymbol{X_i\alpha} + \boldsymbol{C_i\theta} + \epsilon_i, \quad (8)$$

where Y_i is the methylation beta-value for individual i, SNP_i is the number of reference alleles of the SNP, $\boldsymbol{PC_i}$ is a vector of the first four principal components of the genetic relatedness matrix, and $\boldsymbol{X_i}$ is a vector of the control variables age and sex, as well as an interaction term between age and sex. $\boldsymbol{C_i}$ is a vector containing study-specific controls and technical covariates (such as dummy variables for genotyping array and genotyping batches).

9.1 Quality control of methQTL analyses

The GWAS results of the methQTL analyses were quality controlled according to a stringent protocol by the GIANT consortium[56], and the protocol was implemented with the EasyQC software. In summary the following major filters were applied; removal of monomorphic and multi-allelic SNPs, and structural variants such as INDELs; removal of SNPs with an IMPUTE imputation quality < 0.7; and of SNPs with a minor allele frequency (MAF) < 0.05. The quality control procedure ensures that all SNPs have alleles aligned to the 1000G phase 3, version 5 (October 2014 haplotype release)[57], and SNPs that could not be aligned with the reference were removed.

9.2 Results – methQTL analysis

Using the software METAL[29] we performed sample-size weighted meta-analysis across the LBC1936 and RS3 cohorts with a total sample size of 1,649. The meta-analysis resulted in no genome-wide significant ($P < 5\times10^{-8}$) methQTLs for any of the 9 lead probes. We note that the non-significant results could be the result of the relatively small sample size of the meta-analysis, although the current sample size has 80% statistical power to find SNP-effects with R^2 = 2.37% at genome-wide significance. To put these effect sizes into perspective, previous studies find many CpG loci where the variance of the methylation level is explained by up to 89% by SNP effects (Gibbs et al., 2010; Lemire et al., 2015)[55,58]. The observed effect sizes in our methQTL analyses are much smaller (< 2%), and we therefore belive that the magnitude of the possible confounding for the 9 lead probes due to SNP-effects is low compared to SNP-effects observed for other CpGs (that are not associated in our EWAS).

10 References

1. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 2012; **12,** 529–541.

2. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; **14**, 1-19 (2013).

3. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan *et al.* Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease : A Mendelian Randomization Approach. *PLoS Med.* 2017; **14**: 1–30.

4. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* 2016; **98**: 680–696.

5. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016; **17**: 1–15.

6. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* 2016; **9**: 436-447.

7. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson MM, Zhou Y-H *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.* 2015; **24**: 4464–4479.

8. Johnson W, Kyvik KO, Mortensen EL, Batty GD, Deary IJ. Does education confer a culture of healthy behavior? Smoking and drinking patterns in Danish twins. *Am. J. Epidemiol.* 2011; **173**: 55–63.

9. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. [Internet]. 2015 [cited 4 April 2016]. Available from: <https://cran.r-project.org/web/packages/boot/>

10. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; **340**: 1467–1471.

11. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539–542.

12. United Nations Educational, Scientific and Cultural Organization. International Standard Classification of Education [Internet]. 2006 [cited 18 September 2015]. Available from: <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>

13. Triche TJ, Weisenberger DJ, Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* 2013; **41**: 1–11.

14. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 2014; **15**: 503-520.

15. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012; **13**: 86-102.

16. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Lifang H *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010; **11**: 587-596.

17. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P-C, Reynolds LM *et al.* A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* 2016; e-pub ahead of print 15 November 2016; doi:10.1038/mp.2016.192

18. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H *et al.* Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech. Ageing Dev.* 2009; **130**: 234–239.

19. Langevin SM, Houseman EA, Christensen BC, Wiencke JK, Nelson HH, Karagas MR *et al.* The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics* 2011; **6**: 908–919.

20. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MPM, Eijk, K *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012; **13**: R97.

21. Bell JT, Yang TP-C, Pidsley R, Nisbet J, Glass D, Mangino M *et al.* Epigenome-wide

scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 2012; **8**: e1002629.

22. Florath I, Butterbach K, Müller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: An epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum. Mol. Genet.* 2012; **23**: 1186–1201.

23. Boks MP, Derks EM, Weisenberger, DJ, Strengman E, Janson E, Sommer IE *et al.* The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* 2009; **4**: 21–23.

24. Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. *PLoS One* 2010; **5**: e10028.

25. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K *et al.* Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics* 2011; **6**: 623–629.

26. Barro RJ, Lee J-W. International data on educational attainment: updates and implications. *Oxf. Econ. Pap.* 2001; **53**: 541–563.

27. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 2013; **6**: 4-19.

28. Chen Y-A, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013; **8**: 203–209.

29. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**: 2190–2191.

30. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 2011; **19**: 807–812.

31. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

32. Iterson M, Zwet E, the BIOS Consortium, Heijmans BT. Controlling bias and inflation

in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017; **18**: 1–13.

33. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. *Stat. Med.* 2014; **33**: 1946–1978.

34. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539–542.

35. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai P-C *et al.* DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging* 2016; **8**: 1844–1865.

36. Johnson W, Kyvik KO, Mortensen EL, Batty GD, Deary IJ. Does education confer a culture of healthy behavior? Smoking and drinking patterns in Danish twins. *Am. J. Epidemiol.* 2011; **173**: 55–63.

37. Gao X, Jia M, Zhang Y, Breitling LP, Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* 2015; **7**, 113-123.

38. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL *et al.* Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: Findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum. Mol. Genet.* 2015; **24**: 2201–2217.

39. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* 2010; **42**: 441–447.

40. Karlsson Linnér R, Beauchamp JP. Large-scale genetic study of risk tolerance and risky behaviors identifies new loci and reveals shared genetic influences. *Manuscript in preparation for publication* 2017.

41. Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci.* 2014; **111**: 13790–13794.

42. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* 2013; **14**: 507–515.

43. Shah S, Bonder MJ, Marioni, RE, Zhu Z, McRae AF, Zhernakova A *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am. J. Hum. Genet.* 2015; **97**: 75–85.

44. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013; **9**: e1003348.

45. Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge University Press: Cambridge, UK, 1997.

46. The R Core Team. R: A language and environment for statistical computing [Internet]. 2013 [cited 4 April 2016]. Available from: <https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>.

47. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H *et al.* Data Resource Profile : Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int. J. Epidemiol.* 2015; **44**: 1181–1190.

48. GOV.UK. National Curriculum [Internet]. 2016 [cited 14 January 2017]. Available from: <https://www.gov.uk/national-curriculum>

49. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C *et al.* Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS One* 2013; **8**: e63812.50. Fraser, A. *et al.* Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42,** 97–110 (2012).

51. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Erst J, Bilenky M, Yen A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**: 317–330.

52. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 2013; **45**: 580–585.

53. Watanabe K, Taskesen E, Bochoven A, Posthuma D. FUMA: Functional mapping and annotation of genetic associations. *Manuscript submitted for publication* 2017.

54. Quon, G., Lippert, C., Heckerman, D. & Listgarten, J. Patterns of methylation heritability in a genome-wide analysis of four brain regions. *Nucleic Acids Res.* **41,** 2095–2104 (2013).

55. Lemire M, Zaidi SHE, Ban M, Ge B, Aïssi D, Germain M *et al.* Long-range epigenetic

regulation is conferred by genetic variation located at thousands of independent loci. *Nat. Commun.* 2015; **6**: 6326.

56. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* 2014; **9**: 1192–1212.

57. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.

58. Gibbs JR, Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai S-L *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in Human Brain. *PLoS Genet.* 2010; **6**: e1000952.

11 Author contributions

Daniel Benjamin, Philipp Koellinger, Ian Deary, Niels Rietveld, and Riccardo Marioni designed and oversaw the study.

The lead analysts responsible for the EWAS quality control and meta-analyses were Richard Karlsson Linnér and Niels Rietveld.

Riccardo Marioni performed the epigenetic clock analyses, and the results were meta-analysed by Richard Karlsson Linnér and Niels Rietveld.

Niels Rietveld (RS-BIOS and RS3), Riccardo Marioni (LBC1936), and Andrew Simpkin together with Neil Davies (ALSPAC), performed the polygenic prediction.

Richard Karlsson Linnér and Niels Rietveld performed the investigation of possible confounding with tobacco smoking.

Richard Karlsson Linnér and Niels Rietveld designed and performed the enrichment analyses.

Richard Karlsson Linnér designed and performed the analysis of the tissue-specific methylation.

Kyoko Watanabe performed the FUMA expression analysis.

The methQTL GWAS analyses were performed by Riccardo Marioni (LBC1936), and by Niels Rietveld (RS). Richard Karlsson Linnér performed the quality control and meta-analysis.

All authors contributed to and critically reviewed the manuscript. Daniel Benjamin, Richard Karlsson Linnér, Riccardo Marioni, and Niels Rietveld made especially major contributions to writing and editing.

12 Additional acknowledgments

ALSPAC (Avon Longitudinal Study of Parents and Children) – We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This research was specifically funded by the Economics and Social Research Council (ESRC) ES/N000382/1, ES/N000498/1, ES/N000757/1, and The Medical Research Council (MRC) and the University of Bristol fund the MRC Integrative Epidemiology Unit (MC_UU_12013/1, MC_UU_12013/2, MC_UU_12013/9). GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary, available at:

<https://www.bris.ac.uk/aslpac/researchers/data-access/data-dictionary/>

BSGS (Brisbane System Genetics Study) – We gratefully acknowledge the participants and their families. We would like to thank Lisa Bowdler and Steven Crooks for their technical assistance with the microarray hybridizations, Alison Mackenzie, Marlene Grace and Ann Eldridge for data collection and Dale Nyholt and Scott Gordon for data management. We acknowledge funding by Australian National Health and Medical Research Council (NHMRC) grants 241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498 and Australian Research Council grants A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921 for building and maintaining the adolescent twin family resource through which samples were collected. This research was supported by NHMRC grants 389892, 496667 and 613601.

EPIC (The European Prospective Investigation into Cancer and NutritionItaly) – This research was supported by the "Lifepath" grant to Paolo Vineis at Imperial College, London, and Silvia Polidoro at the HuGeF Foundation (European Commission H2020 grant, Proposal number: SEP-210176796). Paolo Vineis and Silvia Polidoro are funded through the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 633666.

EGCUT (Estonian Genome Center, University of Tartu) – This research was funded by Estonian Research Council Grant IUT20-60, EU H2020 grant 692145, and the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) GENTRANSMED. This work was carried out in part in the High Performance Computing Center of University of Tartu.

FTC (Finnish Twin Cohort) – We warmly thank the participating twin pairs and their family members for their contribution. We would like to express our appreciation to the skilled study interviewers A-M Iivonen, K Karhu, H-M Kuha, U Kulmala-Gråhn, M Mantere, K Saanakorpi, M Saarinen, R Sipilä, L Viljanen, and E Voipio. Anja Häppölä and Kauko Heikkilä are acknowledged for their valuable contribution in recruitment, data collection, and

data management. Phenotyping and epigenotyping of the Finnish twin cohorts has been supported by the Academy of Finland Center of Excellence in Complex Disease Genetics grants 213506, 129680; the Academy of Finland grants 100499, 205585, 118555, 141054, 265240, 263278 and 264146 (to prof. Jaakko Kaprio), and 251316 (to dr. Miina Ollikainen), and 277209 (to dr. Antti Latvala); NIH Grant DA12854 (to prof. Pamela A.F. Madden, Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA); Sigrid Juselius Foundation (to prof. Jaakko Kaprio and dr Miina Ollikainen); Helsinki University Research Grants (to dr. Miina Ollikainen); EPITRAIN - FP7-PEOPLE-2012-ITN, Grant Agreement 316758 funded by the European Union's Seventh Framework Programme.

FINRISK (The National FINRISK Study) – DILGOM methylation data: Finnish Academy grant no. 255935 and 269517.

HBCS (Helsinki Birth Cohort Study) – We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation.

KORA (Cooperative Health Research in the Region of Augsburg F4) – The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

LBC (Lothian Birth Cohort) – We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society, and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. Genotyping of the cohorts was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. The raw data collected in the course of our research with human participants is available upon request (please contact, by e-mail, Professor Ian Deary, at the University of Edinburgh, and ask for a 'Lothian Birth Cohort Data Request Form'). The process is facilitated by a full-time Lothian Birth Cohort database manager. Such proposals, when approved, are conducted in collaboration with appropriate members of the Lothian Birth Cohort study team.

LLD (Lifelines Deep) – The LifeLines Cohort Study is a large population-based cohort study and biobank that was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing. Between 2006 and 2013, inhabitants of the northern part of The Netherlands and their families were invited to participate, thereby contributing to a three-generation design. This work was supported by the European Research Council Advanced Grant (ERC-671274 to CW), the Dutch Digestive Diseases Foundation (MLDS WO11-30 to CW), the European Union's Seventh Framework Programme (EU FP7) TANDEM project (HEALTH-

F3-2012-305279 to CW), the Netherlands Organization for Scientific Research (NWO-VENI grant 916-10135 to LF and NWO VIDI grant 917-14374 to LF). Generation of the methylation data (as part of the Biobank-based Integrative Omics Study (BIOS)) is financially supported by the Biobanking and Biomolecular Research Infrastructure of The Netherlands (BBMRI-NL), funded by the Netherlands Organisation for Scientific Research (NWO).

MCCS (Melbourne Collaborative Cohort Study) – This work was supported by the Australian National Health and Medical Research Council (NHMRC, grant 1088405). MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria, and the methylation studies were supported by NHMRC grants 1011618, 1026892, 1027505, 1050198, 1043616 and 1074383.

OATS (Older Australian Twins Study) – We would like to acknowledge and thank the Older Australian Twins Study (OATS) participants and the OATS research team. OATS was supported by a National Health and Medical Research Council (NHMRC)/Australian Research Council Strategic Award (Grant 401162) and the NHMRC Project Grant (1045325). OATS was facilitated through access to the Australian Twin Registry, which is funded by the NHMRC Enabling Grant 310667. In addition, we would like to acknowledge the NHMRC grants 613608 and 613602, which contributed funding for the DNA methylation assays.

RS (Rotterdam Study) – The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk PhD for their help in creating the methylation database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. Some of the analyses were performed on the Dutch national e-infrastructure with the support of SURF Cooperative.

SATSA: Swedish Adoption/Twin Study of Aging – This work was supported by NIH grants R01 AG04563, AG10175, AG028555, the MacArthur Foundation Research Network on Successful Aging, the Swedish Council for Working Life and Social Research (FAS/FORTE) (97:0147:1B, 2009-0795, 2013-2292), the Swedish Research Council (825-2007-7460, 825-2009-6141, 521-2013-8689, 2015-03255), Karolinska Institutet delfinansiering (KID), Loo & Hans Osterman Foundation, the Foundation for Geriatric Diseases and Erik Rönnberg donation.

YFS (The cardiovascular risk in Young Finns Study) – The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio

Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

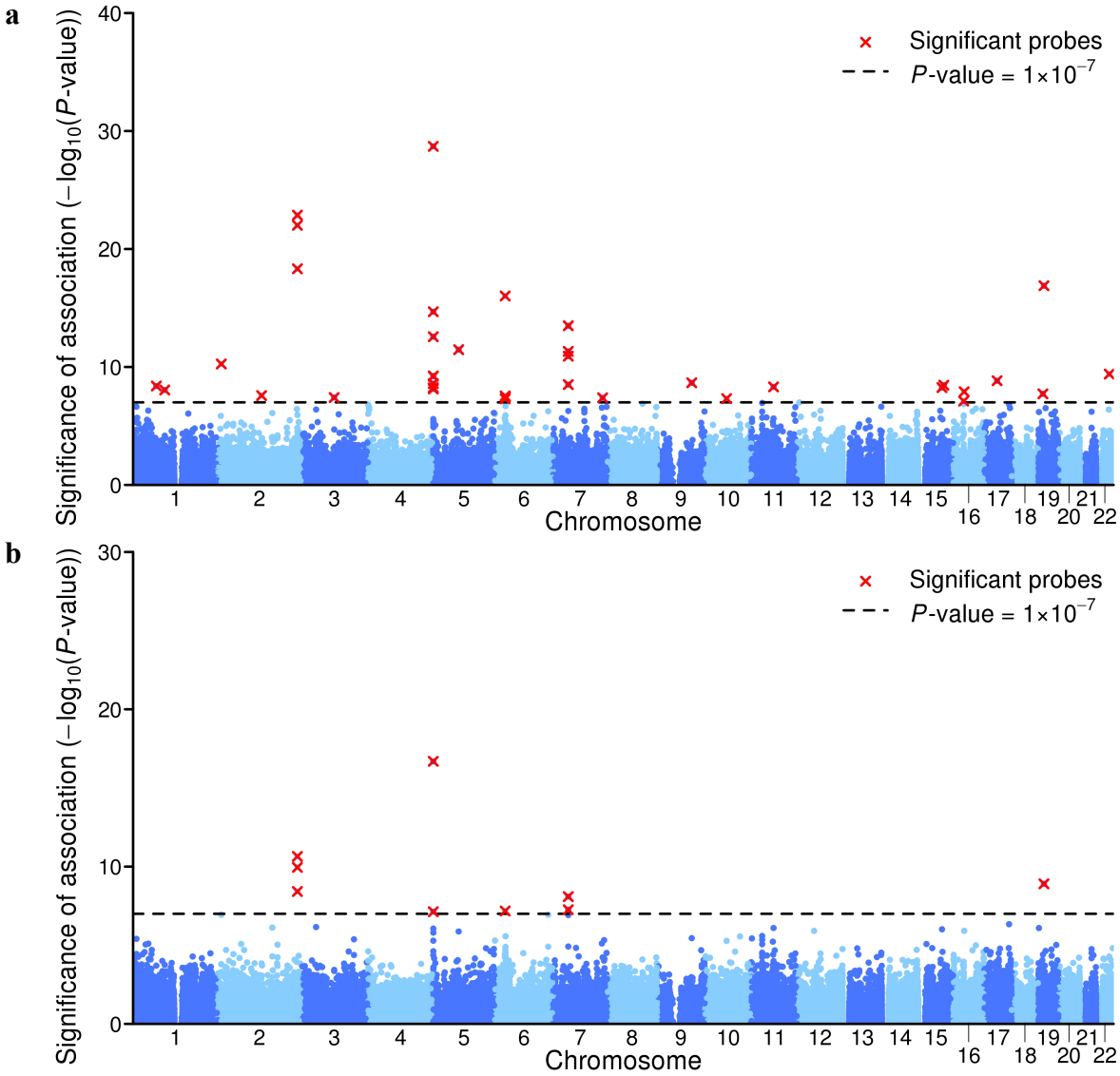
The expert technical assistance in the statistical analyses by Irina Lisinen is gratefully acknowledged.

The YFS has been approved by the 1st ethics committee of the Hospital District of Southwest Finland on December 19th, 2006, January 15th, 2008, and September 21st, 2010.

Individual acknowledgments – Richard Karlsson Linnér gratefully acknowledges SURFsara for providing the computation infrastructure, and funding from the European Research Council through my promotor Philipp Koellinger (ERC consolidator grant 647648 EdGe). Niels Rietveld gratefully acknowledges funding from the Netherlands Organization for Scientific Research (NWO Veni grant 016.165.004). Philipp Koellinger gratefully acknowledges funding from the European Research Council (ERC consolidator grant 647648 EdGe). Danielle Posthuma acknowledges financial support from The Netherlands Organization for Scientific Research (NWO VICI 453-14-005, NWO Complexity 645-000-003). Daniel Benjamin gratefully acknowledges funding from NIA/NIH through grant R01-AG042568-02 to the University of Southern California.

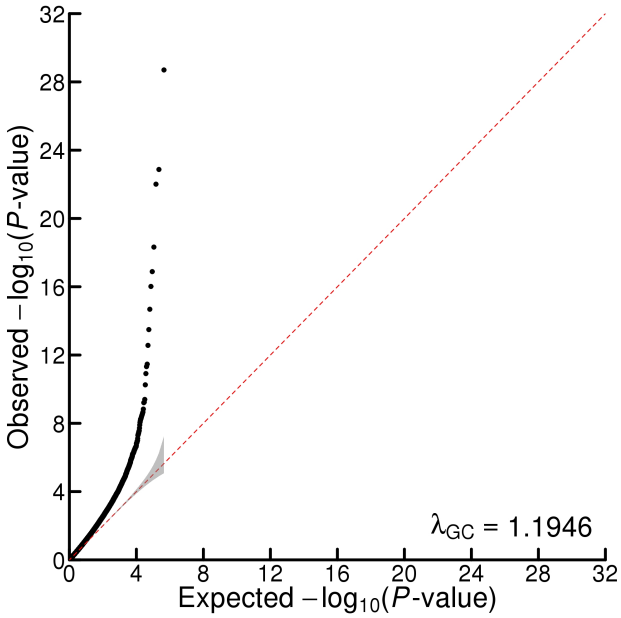
13 Supplementary figures

Supplementary Figure 1 – Manhattan plots of the basic and adjusted EWAS models. Manhattan plots of the EWAS meta-analysis of EA from **(a)** the basic model, and **(b)** the adjusted model. The *x*-axis is chromosomal position, and the *y*-axis is the significance on a $-\log_{10}$ scale. The dashed line marks the threshold for epigenome-wide significance ($P = 1\times10^{-7}$). Each epigenome-wide significant association is marked with a red ×.

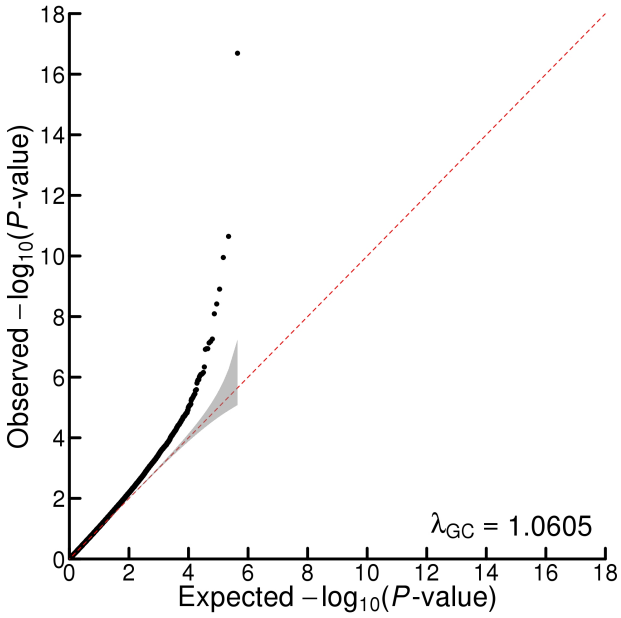


Supplementary Figure 2 – QQ plots of the basic and adjusted EWAS models. Quantile-quantile (QQ) plots of the EWAS meta-analysis of EA for **(a)** the basic model and **(b)** the adjusted model. The gray shaded areas in the QQ plots represent the 95% confidence intervals under the null hypothesis. The presented genomic inflation factor, λ_{GC}, is calculated after probe filtering and cohort-level genomic control.

a

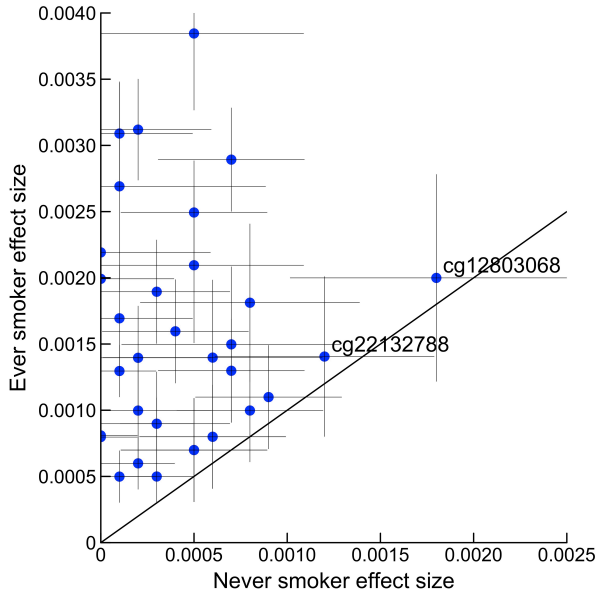


b

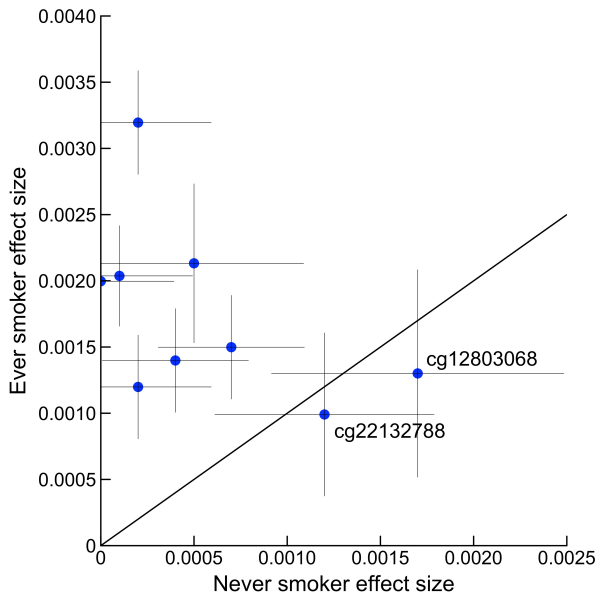


Supplementary Figure 3 – The effect size comparison between ever smokers and never smokers. Effect size estimates with 95% confidence intervals in the ever smokers plotted against the effect size estimates in the never smokers. Panel **(a)** displays the effect size estimates from the meta-analysis of the re-estimation in the never-smoker subsample (*N* = 5,175) for the basic model, and panel **(b)** displays the comparison for the adjusted model. We aligned the effect sizes to the first quadrant by taking their absolute values. The line represents the 45-degree line. The two probes discussed in **Supplementary Note 2.3.3**; cg12803068 and cg22132788, are marked in the figure.

a



b



Supplementary Figure 4 – Heatmap of the expression analysis using GTEx expression data. The heatmap created with the webtool FUMA[53], displays the normalised expression for the 29 genes in proximity with the CpG probes with *P*-value < 1×10^{-5} of the adjusted EWAS model. The color spectrum represents the normalized relative expression across tissues per gene (Reads Per Kilo base per Million, RPKM). Two clusters were observed and these are marked by red lines, the first cluster consists of the 10 genes – *KCNH3, CLSTN1, YWHAQ, LINC00277, C10orf107, RNF220, MKRN3, GAL3ST3, PFZD7*, and *FLRT1*, and this cluster is observed as more highly expressed in brain tissues compared to the other tissues. The cluster is located between "Brain_Spinal_cord_cervical_c-1" and "Brain_Nucleus_accumbens_basal_ganglia" in the top-mid part of the heatmap. The 19 genes *DAPL1, WT1, MYO15B, MYO1G, STK40, ASCL2, AHRR, DOT1L, SSR3, POLK, SP3, ARID1B, ZNF267, ZNF808, C1D, EXOC2, F2RL3, CDCP1,* and *IER3*, form a second cluster of genes that are more highly expressed in a wide variety of non-brain tissues compared to the brain tissues. The second cluster is located in the bottom-left corner between "Cells_EBV-transformed_lymphocytes" and "Uterus". The method is described in **Supplementary Note 8**.